Innovative Systems Design and Engineering
ISSN 2222-1727 (Paper) ISSN 2222-2871 (Online)
Vol 3, No 6, 2012
www.iiste.org

Estimating Markov Model Parameters from Statistical Analysis of Speech Packets Transmitted over BWA Networks

Oleg Sheluhin[1] Aderemi Atayero[2*] Adeyemi Alatishe[2] Juliet Iruemi[2]

1. Dept. of Info. Security, Moscow Technical Univ. of Comm. & Informatics, Russia
2. Dept. of Electrical & Info. Engineering, Covenant University, pmb1023 Ota, Nigeria

* E-mail of the corresponding author: atayero@{ieee.org ; covenantuniversity.edu.ng}

Abstract

The main problem encountered in the design of Markov models is in choosing the adequate mathematical model for their representation. Analysis of the mathematical models of speech packet sources was conducted on the basis of Markov chains. The specific characteristics of telephony dialogue and monologue were taken into consideration. These in their turn can be used in the design of methods of statistical companding of packet communication nodes. In this paper we show that the choice of the number of states of the Markov chain as well as their probabilistic characteristics can be estimated from the results of the statistical processing of speech packets transmitted over Broadband Wireless Access Networks (BWAN).

Keywords: Markov models, Wireless network, Speech processing, Packet switching, Mathematical model.

1. Introduction

The statistical analysis of fragments of speech signal has received a lot of attention in works on compression (Sheluhin et al., 2012; Lajos et al., 2001; Rasool and Sadegh, 2007; Peter and Rainer, 2006) and transmission (Lajos et al., 2001; Peter and Rainer, 2006; Atayero, 2000) of speech reported in the literature. As a result of these investigations, a series of models describing the change in the signal state during a telephone conversation have been proposed (Peter and Rainer, 2006; Minoli and Minoli, 1999; Gringeri et al. 2007). In general, the complexity of the proposed models increases with the level of accuracy of experimental data obtained in the measurement of the parameters of speech signal. Markov processes (MP) with the necessary number of states describe the mechanism of formation of speech signal sufficiently well. This knowledge is necessary for the analysis of network problems during packet speech communication.

Accuracy of the model is determined by the ability to predict the duration of different states of speech signal. The distribution of each of these events can be obtained with necessary accuracy through a majority of known and relatively detailed models (Lajos et al., 2001; Peter and Rainer, 2006; Minoli and Minoli, 1999; Wuncheol et al. 2004; Tanenbaum, 2006; Sangwan et al., 2002)). Less complex models describe the distribution of only certain states, for example: the duration of active speech state or pause duration. However, these states are the most important for the analysis of traffic models and network statistical characteristics. Analytic expressions convenient for the analysis of major network characteristics can be obtained via simplified models, which can be used for the analysis of network with statistical companding.

The result of the statistical analysis of speech signal is the design of models describing the change in state of speech in the course of a telephone conversation. We enumerate the basic steps usually executed in the design of models of speech processes:

1. The first stage entails choosing the mathematical apparatus for the description of real processes;
2. The second stage entails checking of the statistical hypotheses on the suitability of the models to real processes, to which end the correspondence criterion needs be chosen;
3. At the third stage, numerical values of the parameters are chosen and the main characteristics of investigated processes are computed;
4. The fourth stage is concerned with the use of derived models for the analysis of concrete technical systems.

In practice, the second stage is often carried out simultaneously with the third or fourth.

1.1. Problem Statement

The main problem encountered in the design of Markov models is in choosing the adequate mathematical model. This problem converges in the resolution of two tasks:

1. Justification of the choice of number of states of the Markov chain (MC);
2. Adequate estimation of the transition probability parameters of the MC from one state to the other using the results of the statistical processing of speech signals.

This paper is devoted to the resolution of these tasks. Two main types of Markov models are analyzed – Dialogue models and Monologue models. G728 and G711 codecs were used as sources of speech packets.

2. Dialogue Models

The research problem of studying speech dialogue dynamics has been identified for quite a while, and quite a lot of scientific work has been published on it. Assume the existence of a speech transmission device in which each subscriber is given a certain time interval for the transmission of either a full or empty packet, depending on the speech signal energy level. The packet flow is not synchronized with moments of occurrence of active speech signal. We assume henceforth that only one of the subscribers engaged in a dialogue can change the state of speech signal within the interval of one packet (Sheluhin et al., 2012).

2.1 The *Brody* Six-state Model

This model is made up of six possible states in which speech can be at any given time in a dialogue:

- State 1: subscriber A is speaking, subscriber B is silent;
- State 2: subscriber A is preparing to enter *pause* state;
- State 3: both subscribers are silent, subscriber A spoke last;
- State 4: subscriber B is speaking, subscriber A is silent;
- State 5: subscriber B is preparing to be silent;
- State 6: both subscribers are silent, subscriber B spoke last.

The *Brody* model adequately describes the dynamics of all possible dialogue states. A graph of this model is presented on Figure. 1.

The corresponding probability matrix for this model is as given in equation (1).

$$P = \begin{Vmatrix} p_{11} & p_{12} & 0 & p_{14} & 0 & 0 \\ p_{21} & p_{22} & 0 & 0 & 0 & p_{26} \\ p_{31} & 0 & p_{33} & 0 & 0 & p_{36} \\ p_{41} & 0 & 0 & p_{44} & 0 & p_{46} \\ p_{51} & 0 & 0 & 0 & p_{55} & p_{56} \\ 0 & 0 & p_{63} & 0 & p_{65} & p_{66} \end{Vmatrix} \qquad (1)$$

2.2 The Four-state Model[1]

This model comprises of the following states:

- State 1: subscriber A is speaking, subscriber B is silent;
- State 2: subscriber A is speaking, subscriber B is speaking;
- State 3: subscriber A is silent, subscriber B is speaking;
- State 4: subscriber A is silent, subscriber B is silent.

The graph of this model is depicted on Figure. 2, and its corresponding state probability matrix is as given in equation (2).

[1](Minoli and Minoli, 1999)

$$P = \begin{Vmatrix} p_{11} & p_{12} & p_{13} & 0 \\ p_{21} & p_{22} & 0 & p_{24} \\ p_{31} & 0 & p_{33} & p_{34} \\ 0 & p_{42} & p_{43} & p_{44} \end{Vmatrix} \tag{2}$$

This model adequately depicts distribution of the duration of *pause* and *active speech*; it however does not reflect real event flow in cases of simultaneous speech activity by both subscribers.

2.3 The Three-state Model

The three-state model is made up of the following states:

- State 1: subscriber A is speaking, subscriber B is silent;
- State 2: subscriber A is silent, subscriber B is silent;
- State 3: subscriber A is silent, subscriber B is speaking.

The three-state model is derived by excluding the possibility of simultaneous speech activity by both subscribers, which corresponds to State 2 of the four-state model. The graph of the three-state model and its corresponding state probability matrix are given in Figure. 3 and equation (3) respectively.

$$P = \begin{Vmatrix} p_{11} & p_{12} & 0 \\ p_{21} & p_{22} & p_{23} \\ 0 & p_{32} & p_{33} \end{Vmatrix} = \begin{Vmatrix} 1-p & p & 0 \\ x & 1-x-y & y \\ 0 & y & 1-r \end{Vmatrix} \tag{3}$$

2.4 Two-state model

The simplest model of speech dialogue achievable is the two-state model, which is derived by removing the second state of the three-state model. It consists of the following two states:

- State 1: subscriber A is speaking, subscriber B is silent;
- State 2: subscriber A is silent, subscriber B is speaking;

This is a well-researched model, that allows for obtaining analytic results. Its graph is presented in Figure. 4, and its corresponding state probability matrix is as given in equation (4). It is majorly employed for modeling channels and networks used for packet transmission of speech signal.

$$P = \begin{Vmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{Vmatrix} = \begin{Vmatrix} 1-p & p \\ r & 1-r \end{Vmatrix} \tag{4}$$

The 2-state model corresponds to experimental data on the distribution of the active speech phase. It however does not reflect the peculiarities of the distribution of pause duration. Since the model excludes the possibility of both subscribers being in the pause state simultaneously, the dialogue activity seems to be switched between subscribers, while both the active and pause states have a geometric distribution.

We create our model based on the preceding condition, while taking into consideration the fact that in the course of a dialogue, a subscriber is either in control of the dialogue or responding in reaction to information received from the other subscriber. In this case the model will consist of two main states (active and passive), with the possibility of having sub-states, which allow for short-term transition into active state inside a pause or passive state in active state (Sheluhin et al., 2012).

The models presented above show that the subjectivity of designing possible mathematical models of speech dialogues. The approach of building mathematical models based on the analysis of objective statistical characteristics of speech dialogue elicits more interest.

3. Estimation of Dialogue Model Parameters from Empirical Results

- An experiment was conducted to study the properties of speech signal during dialogues between two subscribers. The experimental setup described in details in a previous work by the authors in (Sheluhin et al., 2012) and is hereby summarized: A PC and a PSTN telephone set were used to record telephone dialogues with the aid of a special software;

- The experiment was transparent to the subscribers for authenticity of data;
- The recorded speech was digitized at 8 kHz and 8 bit, saved as PCM .wav files;
- Subscriber active speech and pause in dialogue periods were fixated.

Table 1. gives the results of analyzing a total of ten (10) dialogues.

A comparative analysis of telephone dialogue data and those given in (Peter and Rainer, 2006; Atayero, 2000) suggests that the derived statistical data is a realistic representation of the components of telephone conversation. Hence the possibility of developing a model of the original speech signal based on obtained data vis-à-vis the model of telephone dialogue for subsequent investigation of telephone traffic formation.

The selection of the appropriate analytical expression for description of real processes is required in the process of developing a model of the source of speech packets. It is assumed *ab initio* at the modeling stage that a subscriber's speech can be divided into two distinct states namely: Active (A) and Passive (P). Figure. 5. and Figure. 6. depict relationship F(X > x), corresponding to the probability distributions of active and passive packets respectively.

The experimental characteristics show that the model, while presupposing the presence of two states – A и P, does not reflect the dynamics of change of speech signal with enough exactitude, since the presence of one state A during speech period and one state P in the pause period suggests that the relationship F(X > x) obeys an exponential law.

Further analysis of the distribution function F(X > x) at the next stage revealed that it is accurately approximated by a function of the form:

$$F\left(X>x\right)=S_1\mathrm{e}^{-\alpha_1 x}+S_2\mathrm{e}^{-\alpha_2 x}+S_3\mathrm{e}^{-\alpha_3 x} \tag{5}$$

The distribution function of the active packet series is of the form:

$$F\left(X>x\right)=A_1\mathrm{e}^{-\alpha_1 x}+A_2\mathrm{e}^{-\alpha_2 x}+A_3\mathrm{e}^{-\alpha_3 x} \tag{6}$$

While the distribution function of the passive packet series is:

$$F\left(X>x\right)=P_1\mathrm{e}^{-\beta_1 x}+P_2\mathrm{e}^{-\beta_2 x}+P_3\mathrm{e}^{-\beta_3 x} \tag{7}$$

The generic form of (2) and (3) is given as (1), in which S denotes state. Eq. (2) is the distribution of conditional probabilities of the occurrence of a series of Active states conditioned on the previous state being Passive. However, Eq. (3) is the conditional probability distribution of the occurrence of a series of Passive states conditioned on the previous state being Active.

Tables 2 and 3 present approximations of the distribution function gotten for the active and passive speech packet series respectively.

The type of experimental characteristic F(X > x) and their representation lead to the logical assumption of the existence of three states each corresponding to active − A1, A2 and A3, as well as pause in speech − P1, P2 and P3 respectively, with different expectation values for active and passive states' duration. The developed model has a graph of the form shown in Figure. 7 (Sheluhin et al., 2012).

For this model, the matrix of the transition probabilities from state to the state has the form given in equation (5).

$$S_{ij} = \begin{pmatrix} S_{A_1A_1} & 0 & 0 & S_{A_1P_1} & S_{A_1P_2} & S_{A_1P_3} \\ 0 & S_{A_2A_2} & 0 & S_{A_2P_1} & S_{A_2P_3} & S_{A_2P_3} \\ 0 & 0 & S_{A_3A_3} & S_{A_3P_1} & S_{A_3P_2} & S_{A_3P_3} \\ S_{P_1A_1} & S_{P_1A_2} & S_{PA} & S_{P_1P_1} & 0 & 0 \\ S_{P_2A_1} & S_{P_2A_2} & S_{PA} & 0 & S_{P_2P_2} & 0 \\ S_{P_3A_1} & S_{P_3A_3} & S_{PA} & 0 & 0 & S_{P_3P_3} \end{pmatrix} \tag{8}$$

From mere analysis of the distribution function of the length of a series of packets obtained from experimental data, It is impossible to determine to which of the three states a particular packet or series of packets belongs. The same holds for states P1, P2 or P3 – it is impossible to determine to which of the states belongs a particular packet or series of packets, making up a given period of pause.

It ergo becomes subsequently necessary to find a compromise in the process of dividing one state into several *substates*. A likely resolution of this problem emanating from the condition of preservation of the final probabilities of signal division into pause and active speech states is considered. Equation (6) is introduced for ease of expression:

$$S_{ij} = Q_{ij} = \begin{pmatrix} q_{11} & 0 & 0 & q_{14} & q_{15} & q_{16} \\ 0 & q_{22} & 0 & q_{24} & q_{25} & q_{26} \\ 0 & 0 & q_{33} & q_{34} & q_{35} & q_{36} \\ q_{41} & q_{42} & q_{43} & q_{44} & 0 & 0 \\ q_{51} & q_{52} & q_{53} & 0 & q_{55} & 0 \\ q_{61} & q_{62} & q_{63} & 0 & 0 & q_{66} \end{pmatrix} \tag{9}$$

Matix Q_{ij} defines the final probabilities.

$$Q_i = [Q_1 \quad Q_2 \quad Q_3 \quad Q_4 \quad Q_5 \quad Q_6]^{\mathrm{T}} \tag{10}$$

Similarly, the graph of the six-state model shown in Figure. 7 can be reduced to that of the four-state model as shown in Figure. 8. For such a model, the transition probabilities matrix from state to state P_{ij} is given as follows:

$$P_{ij} = \begin{array}{c} \begin{array}{cccc} A & P_1 & P_2 & P_3 \end{array} \\ \begin{pmatrix} P_{AA} & P_{AP_1} & P_{AP_2} & P_{AP_3} \\ P_{P_1A} & P_{P_1P_1} & 0 & 0 \\ P_{P_2A} & 0 & P_{P_2P_2} & 0 \\ P_{P_3A} & 0 & 0 & P_{P_3P_3} \end{pmatrix} \end{array} \tag{11}$$

We introduce the matrix in equation (8) for ease of expression:

$$P_{ij} = R_{ij} = \begin{pmatrix} r_{11} & r_{12} & r_{13} & r_{14} \\ r_{21} & r_{22} & 0 & 0 \\ r_{31} & 0 & r_{33} & 0 \\ r_{41} & 0 & 0 & r_{44} \end{pmatrix}. \tag{12}$$

For *x=1, x=2, x=3*, equation (3) becomes

$$F = \begin{cases} P_{AP}P_1\mathrm{e}^{-\beta_1} + \\ +P_{AP}P_2\mathrm{e}^{-\beta_2} + P_{AP}P_3\mathrm{e}^{-\beta_3} & , \quad x = 1 \\ P_{AP}P_1\mathrm{e}^{-\beta_1}\mathrm{e}^{-\beta_1} + \\ +P_{AP}P_2\mathrm{e}^{-\beta_2}\mathrm{e}^{-\beta_2} + P_{AP}P_3\mathrm{e}^{-\beta_3}\mathrm{e}^{-\beta_3} & , \quad x = 2 \\ P_{AP}P_1\mathrm{e}^{-\beta_1}\mathrm{e}^{-\beta_1}\mathrm{e}^{-\beta_1} + \\ +P_{AP}P_2\mathrm{e}^{-\beta_2}\mathrm{e}^{-\beta_2}\mathrm{e}^{-\beta_2} + P_{AP}P_3\mathrm{e}^{-\beta_3}\mathrm{e}^{-\beta_3}\mathrm{e}^{-\beta_3} & , \quad x = 3 \end{cases} \tag{13}$$

Probabilities for the graph shown in Figure. 4 can be calculated from equation (9):

$$F = \begin{cases} r_{12}r_{22} + r_{13}r_{33} + r_{14}r_{44} & , \quad x = 1 \\ r_{12}r_{22}r_{22} + r_{13}r_{33}r_{33} + r_{14}r_{44}r_{44} & , \quad x = 2 \\ r_{12}r_{22}r_{22}r_{22} + r_{13}r_{33}r_{33}r_{33} + r_{14}r_{44}r_{44}r_{44} & , \quad x = 3 \end{cases} \tag{14}$$

From the comparison of equations (9) and (10), we may write the following:

$$r_{ij} = \mathrm{e}^{-\beta_i}, \; r_{12} = P_{AP}P_1 \; ,..., \; r_{14} = P_{AP}P_3 \tag{15}$$

and consequently, the matrix Pij may be rewritten in the following form:

$$P_{ij} = \begin{pmatrix} P_{AA} & P_{AP}P_1 & P_{AP}P_2 & P_{AP}P_3 \\ 1-\mathrm{e}^{-\beta_1} & \mathrm{e}^{-\beta_1} & 0 & 0 \\ 1-\mathrm{e}^{-\beta_2} & 0 & \mathrm{e}^{-\beta_2} & 0 \\ 1-\mathrm{e}^{-\beta_3} & 0 & 0 & \mathrm{e}^{-\beta_3} \end{pmatrix} \tag{16}$$

The accuracy of the model can be improved by introducing three states − A1, A2, A3 − and a state P. Consequently, the transition probabilities matrix Sij of the states speech packet source model *vis-a-vis* the peculiarities of a six-state dialogue will be:

$$S_{ij} = \begin{vmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{vmatrix} \tag{17}$$

where

$$\mathbf{A} = \begin{vmatrix} \mathrm{e}^{-\alpha_1} & 0 & 0 \\ 0 & \mathrm{e}^{-\alpha_2} & 0 \\ 0 & 0 & \mathrm{e}^{-\alpha_3} \end{vmatrix} \tag{17a}$$

$$\mathbf{B} = \begin{vmatrix} (1-\mathrm{e}^{-\alpha_1})P_1 & (1-\mathrm{e}^{-\alpha_1})P_2 & (1-\mathrm{e}^{-\alpha_1})P_3 \\ (1-\mathrm{e}^{-\alpha_2})P_1 & (1-\mathrm{e}^{-\alpha_2})P_2 & (1-\mathrm{e}^{-\alpha_2})P_3 \\ (1-\mathrm{e}^{-\alpha_3})P_1 & (1-\mathrm{e}^{-\alpha_3})P_2 & (1-\mathrm{e}^{-\alpha_3})P_3 \end{vmatrix} \tag{17b}$$

$$\mathbf{C} = \begin{vmatrix} (1-\mathrm{e}^{-\beta_1})A_1 & (1-\mathrm{e}^{-\beta_1})A_2 & (1-\mathrm{e}^{-\beta_1})A_3 \\ (1-\mathrm{e}^{-\beta_2})A_1 & (1-\mathrm{e}^{-\beta_2})A_2 & (1-\mathrm{e}^{-\beta_2})A_3 \\ (1-\mathrm{e}^{-\beta_3})A_1 & (1-\mathrm{e}^{-\beta_3})A_2 & (1-\mathrm{e}^{-\beta_3})A_3 \end{vmatrix} \tag{17c}$$

$$\mathbf{D} = \begin{vmatrix} \mathrm{e}^{-\beta_1} & 0 & 0 \\ 0 & \mathrm{e}^{-\beta_2} & 0 \\ 0 & 0 & \mathrm{e}^{-\beta_3} \end{vmatrix} \tag{17d}$$

Substituting the numerical values in equation (17) for approximation coefficients earlier obtained (see Table 2.), matrices A, B, C, D become:

$$\mathbf{A} = \begin{vmatrix} 0{,}79453 & 0 & 0 \\ 0 & 0{,}94904 & 0 \\ 0 & 0 & 0{,}99164 \end{vmatrix} \quad (18a) \qquad ; \qquad B = \begin{vmatrix} 0{,}01765 & 0{,}0277 & 0{,}00127 \\ 0{,}04377 & 0{,}00687 & 0{,}00032 \\ 0{,}00719 & 0{,}00113 & 0{,}00005 \end{vmatrix} \quad (18b)$$

$$\mathbf{C} = \begin{vmatrix} 0{,}16166 & 0{,}04876 & 0{,}00922 \\ 0{,}01204 & 0{,}00363 & 0{,}00069 \\ 0{,}00062 & 0{,}00019 & 0{,}00004 \end{vmatrix} \quad (18c) \qquad ; \qquad \mathbf{D} = \begin{vmatrix} 0{,}78036 & 0 & 0 \\ 0 & 0{,}98364 & 0 \\ 0 & 0 & 0{,}999916 \end{vmatrix} \quad (18d)$$

The final probabilities matrix Q_i of the occurrence of the speech packets source in each of the states at any given moment of time can be gotten from matrices 18a–18d:

$$Q_i = \begin{pmatrix} 0{,}110555 \\ 0{,}134454 \\ 0{,}155010 \\ 0{,}120709 \\ 0{,}254354 \\ 0{,}224918 \end{pmatrix} \quad (19)$$

4. Speech Monologue Models

We consider as an example the 3-state transition model the graph of which is presented in Figure. 9. The model describes the three states of speech signal namely: Pause (P), Vocalized speech (V), and Non-vocalized speech (N) (Minoli and Minoli, 1999). The corresponding matrices of transition probabilities (P) and final probabilities (R) for the case depicted on Figure. 9 have the following form:

$$P = \begin{Vmatrix} P_{pp} & P_{pv} & P_{pn} \\ P_{vp} & P_{vv} & P_{vn} \\ P_{np} & P_{nv} & P_{nn} \end{Vmatrix} \quad (20) \qquad ; \qquad R = \begin{Vmatrix} P_p \\ P_v \\ P_n \end{Vmatrix} \quad (21)$$

The final state probabilities are gotten from the system of equations in (22)

$$\textstyle\sum_{i=0}^{K-1} P_i = 1 \quad ; \quad \sum_{j=1}^{K-1} P_j \cdot P_{jk} = P_k \quad (22)$$

4.1 Estimation of Monologue Model Parameters from Experimental Result

As a check of the adequacy of the chosen 3-state monologue Markov model we determine the probability distribution of the length of interval speech signal in pauses (P), Vocalized speech (V) and Non-vocalized speech (N) states. Analysis of a series of speech signals from K windows of 128 samples shows that the length of speech interval in P, V, and N states changes depending on the temp and peculiarities of speech of each individual speaker (Peter and Rainer, 2006; Atayero, 2000). The probability distribution of the length of interval of each of the isolated states is essentially the probability distribution of a discrete random variable $F(m)$, defined as:

$$F(m) = P(M \geq m) \quad (23)$$

where M – random variable, m – a certain value of M that characterizes the number of windows of P (or V or N) in the interval. Histograms are usually constructed to depict graphically the empirical distribution. Data on sample size used in the construction of the probability distribution histogram of length of P, V, N intervals for three (3) Male (M) and two (2) Female (F) speakers is presented in Table 4.

Analysis of the distribution type allowed for suggesting an approximation of the probability distribution function (*pdf*) of interval lengths of the three states of speech considered in the adopted model (P, V, N), in the form of a hyper-exponential expression (Peter and Rainer, 2006).

$$\hat{F}(m) = \sum_i^I A_i\, e^{\alpha_i(m-1)} \qquad ,m = 1, 2, .., \tag{24}$$

where A_i , α_i – coefficients; *m* – number of windows;

$$\sum_i^I A_i = 1 \tag{25}$$

From analysis of results of conducted experiment, it is evident that:

- The distribution of probabilities of P state interval length is well approximated by a weighted sum of two components (I=2), while
- the distribution of probabilities of interval lengths of both the V and N states is adequately approximated by exponential expression with a single component.

In a bid to estimate the quality (accuracy) of approximation, it is necessary to check by means of a statistical criterion the statistical correspondence of experimental distribution with the approximation. By comparing the empirical data with the tabulated statistical values of correspondence criteria, a conclusion can be reached on the level of divergence between the hypothetical function and its estimate. For the distribution of probabilities of interval lengths of states P, V, N, the level closeness of approximating distribution $\hat{F}(m)$ to the empirical $F(m)$ was measured using Kolmogorov criterion (Peter and Rainer, 2006). The maximum divergence of the approximating distribution $\hat{F}(m)$ from the empirical $F(m)$ is computed from equation (26).

$$H_{max} = \max\left|F(m_j) - \hat{F}(m_j)\right| \tag{26}$$

where $j = 1, \dots, n$; *n* – number of experimental points.

Coefficient μ, characterizing the level of importance is calculated from (27)

$$\mu = H_{max}\sqrt{n} \tag{27}$$

By adopting the approach in (Anurag et al., 2004) for the given value of μ we find $P(\mu)$ – the probability that the approximating function was accurately chosen. Data on A_i, α_i, *n*, H_{max}, μ, and $P(\mu)$ for each speaker in one of the states of speech signal *P, V, N* are presented on Tables 5, 6, 7. The average values of the parameters of the approximating functions are also given. The listed values of $P(\mu)$ show that the distribution of the probability of the length of the series that is not less than T for the *V* and *N* states is well approximated by the exponential law, and by the hyper-exponential law for the *P* state. The type of experimental characteristics $F(m)$ and the expressions approximating them allow for assuming the presence of two states, corresponding to pauses $P1$ and $P2$ with different expectation values for the length of the intervals.

4.2 Transition Probability Matrices for Window Flux of Type P, V, N

For the objective of obtaining the statistical characteristics depicting the dynamics of change of speech signal state of the received flux of P, V, N windows, the probability of transition from one state to the other was determined as well as the final probabilities. The final probabilities estimate the continuity of the speech signal in each of the states (Peter and Rainer, 2006; Minoli and Minoli, 1999). The probability transition matrices P_{ij} from state to state and the values for the final probabilities (matrix R) for the speech signals of five speakers M and F were determined.

$$P_{ij} = \begin{Vmatrix} P_{pp} & P_{pv} & P_{pn} \\ P_{vp} & P_{vv} & P_{vn} \\ P_{np} & P_{nv} & P_{nn} \end{Vmatrix} \quad ; \quad R = \begin{Vmatrix} P_p \\ P_v \\ P_n \end{Vmatrix} \tag{28}$$

Thus for example, the elements of the P_{ij} and R matrices for speaker M2 have the following values:

$$P_{ij}^{M2} = \begin{Vmatrix} 0.83980 & 0.01224 & 0.14796 \\ 0.00124 & 0.94153 & 0.05723 \\ 0.19469 & 0.16435 & 0.64096 \end{Vmatrix} \quad ; \quad R = \begin{Vmatrix} 0.23333 \\ 0.57834 \\ 0.18833 \end{Vmatrix} \tag{29}$$

Correspondingly for speaker F1, we have:

$$P_{ij}^{F1} = \begin{Vmatrix} 0.87573 & 0.00731 & 0.11696 \\ 0.00408 & 0.93291 & 0.06301 \\ 0.25559 & 0.22204 & 0.52237 \end{Vmatrix} \quad ; \quad R = \begin{Vmatrix} 0.32548 \\ 0.53547 \\ 0.14905 \end{Vmatrix} \tag{30}$$

An analysis of the transition probabilities reveals that they have a tendency of grouping states, which is evident from the fact that the highest values of the matrix elements are located on the diagonal. It may the concluded that the flux of windows has a complex character, since it is not the algebraic sum of the flux of independent events. The P_{ij} and R matrices were obtained by processing a considerable amount of experimental data. In conjunction with the distributions of the length of series of interval $F(\dot{m})$ given above and the instantaneous signal value, for each state, it may serve as experimental basis for the design of a mathematical model of speech signal. It may be asserted that the Markov model with three states P, V, and N does not fully and adequately describe real speech monologue, since it becomes necessary to subject the pause state to decomposition and determine the elements of the transition probability matrix for the more accurate model with four states P1, P2, V and N (31).

$$P_{ij} = \begin{array}{c} \begin{array}{cccc} P1 & P2 & V & N \end{array} \\ \begin{vmatrix} e^{-\alpha_i} & 0 & (1-e^{\alpha_1}) \cdot K_3 & (1-e^{\alpha_1}) \cdot (1-K_3) \\ 0 & e^{-\alpha_2} & (1-e^{\alpha_2}) \cdot K_3 & (1-e^{\alpha_2}) \cdot (1-K_3) \\ \frac{q_{21} - P_{NP} \cdot A_1 \cdot K_1}{K_1} & \frac{q_{22} - P_{NP} \cdot A_2 \cdot K_2}{K_1} & P_{VV} & P_{VN} \\ \frac{q_{21} - P_{VP} \cdot A_1 \cdot K_1}{K_2} & \frac{q_{22} - P_{VP} \cdot A_2 \cdot K_1}{K_2} & P_{NV} & P_{NN} \end{vmatrix} \end{array} \tag{31}$$

where $K1 = \frac{P_V}{P_V + P_N}$; $K2 = 1 - K1$

As an example, for speaker M2 equation (32) defines the following transition probability matrix for the more accurate model:

$$P_{ij}^{M2} = \begin{array}{c} \begin{array}{cccc} P1 & P2 & V & N \end{array} \\ \begin{vmatrix} 0.73345 & 0 & 0.02037 & 0.24618 \\ 0 & 0.93895 & 0.05639 & 0.00466 \\ 0.00099 & 0.00024 & 0.94154 & 0.05723 \\ 0.15575 & 0.03894 & 0.16435 & 0.64096 \end{vmatrix} \end{array} \tag{32}$$

Depending on the required accuracy for describing the dynamics of change of state of speech signal, the matrix given in expression (31) for the 4-state model may be employed.

Conclusion

We have presented in this paper a method for estimating the parameters of Markov models by analysis of the statistical characteristics of speech transmitted over a wireless conduit. The paper considered various existing models for describing the change in speech signal state during a telephone conversation. It was established that the complexity of models increases with the level of accuracy of experimental data obtained in the measurement of speech parameters. Expressions convenient for the analysis of major network characteristics obtainable via simplified models, which can be used for the analysis of networks with statistical companding of channels, were presented. Detailed procedures for estimating the parameters of Markov models of both speech monologue and dialogue from empirical data was presented. The developed matrices can be used (for example) for analyzing a system of statistical companding. By augmenting the model of probabilities distribution of instantaneous signal values for each state with the result of analysis of experimental readings of real speech signal, we can obtain a sufficiently accurate tool for modeling different types of communication systems. Thus providing a methodology for resolving a main problem encountered in the design of Markov models i.e. that of choosing the adequate mathematical model for their representation.

References

Anurag Kumar, et al. (2004), "Communication Networking: An Analytical Approach", Morgan Kaufman Publishing, San Francisco, USA.

Atayero A.A. (2000), "Estimation of the Quality of Digitally Transmitted Analogue Signals over Corporate VSAT Networks", (Ph.D Thesis).

Gringeri Steve et al. (2007), "Optimizing Transport Services to Integrate TDM and Packet Services", National Fiber Optic Engineers Conference-Networking Convergence for Multi-Haul Architecture, Anaheim, CA.

Lajos Yanzo et al. (2001), Voice compression and communications Principles and Applications for Fixed and Wireless Channels. Willey Interscience. 2001.

Minoli D., Minoli E., (1999) "Delivering Voice over IP Networks", John Willey & Sons Inc., 1999.

Peter Vary and Rainer Martin (2006), "Digital speech transmission: Enhancement, Coding & Error concealment, John Wiley & Sons.

Rasool Tahmasbi and Sadegh Rezaci (2007), "A Soft Voice Activity Detection Using GARCH Filter and Variance Gamma Distribution", IEEE Transactions on Audio, Speech, and Language Processing, Vol. 15, No. 4, pp. 1129-1134.

Sangwan, A. et al. (2002), "VAD Techniques for Real-Time Speech Transmission on the Internet", 5th IEEE International Conference on High-Speed Networks and Multimedia Communications, pp.46~50.

Sheluhin, O.I. et al. (2012), Markov models of telephone speech dialogue, International Journal of Engineering and Technology, vol.2, No.3, pp.328-333, Mar. 2012

Tanenbaum Andrew (2006), "Computer Networks", Prentice-Hall of India Private, New Delhi, India.

Wuncheol Jeong et al. (2004), "Spectral Efficiency of Dynamic Time-Division System for Asymmetric and Dynamic Multimedia Traffic, International Journal of Wireless Information Networks", Vol. 11, No. 4, pp.173~185.

Table 1. Result of Audio Files Analysis

Conversation №	№ of LD-CELP format samples	№ of active samples	№ of passive samples	Quantity of useful information in conversation, % of total volume
01.	62720	18317	44403	29
02.	781622	233763	547859	29
03.	147030	50458	96572	34
04.	529078	168459	360619	31
05.	292284	155586	136698	53
06.	132233	62691	69542	47
07.	917907	254628	663279	27
08.	200608	83336	117272	41
09.	25225	5987	19238	23
10.	1054297	322948	731349	30
Total №	4143004	1356173	2786831	32

Table 2. Approximation coefficients for distribution function of active packet series

Coefficient	A_1	A_2	A_3	α_1	α_2	α_3
Value	0.736	0.222	0.043	0.23	0.0523	0.0084

Table 3. Approximation coefficients for distribution function of passive packet series

Coefficient	P_1	P_2	P_3	β_1	β_2	β_3
Value	0.859	0.1348	0.0065	0.248	0.0165	0.00084

Table 4. Sample size used for plotting the interval length distribution for *P, V, N* windows

SPEAKER	*P*	*V*	*N*
M1	1,192	2,175	839
M2	980	2,429	791
M3	1,084	2,161	955
W1	1,367	2,207	626
W2	978	2,340	882

Table 5. Probability Distribution of P Interval Length

SPEAKER (f_S)	VALUE OF PARAMETERS							
	A_1	α_1	A_2	α_2	n	H_{max}	λ	$P(\lambda)$
M1(8 kHz)	0.80	0.25	0.20	0.058	50	0.032	0.23	≈1
M2(8 kHz)	0.80	0.31	0.20	0.063	45	0.082	0.55	0.9228
M3(8 kHz)	0.70	0.52	0.30	0.048	50	0.080	0.56	0.9228
W1(8 kHz)	0.80	0.21	0.20	0.048	50	0.080	0.56	0.9228
W2(8 kHz)	0.80	0.29	0.20	0.088	45	0.067	0.45	0.9874
M4(32 kHz)	0.64	0.35	0.36	0.083	50	0.060	0.42	0.9874
W3(32 kHz)	0.67	0.37	0.33	0.090	50	0.070	0.49	0.9639
AVERAGE VALUES	0.744	0.329	0.256	0.0683	48.6	0.0673	0.466	0.95118

Table 6. Probability Distribution of V Interval Length

SPEAKER (f_S)	VALUE OF PARAMETERS					
	A	α	n	H_{max}	μ	$P(\mu)$
M1(8 kHz)	1	0.076	58	0.065	0.50	0.9639
M2(8 kHz)	1	0.054	65	0.067	0.54	0.9228
M3(8 kHz)	1	0.088	60	0.049	0.38	0.9972
W1(8 kHz)	1	0.070	57	0.037	0.29	≈1
W2(8 kHz)	1	0.083	55	0.073	0.54	0.9228
M4(32 kHz)	1	0.045	60	0.048	0.37	0.9972
W3(32 kHz)	1	0.048	58	0.050	0.38	0.9972
AVERAGE VALUES	1	0.0663	59	0.0556	0.429	0.97159

Table 7. Probability Distribution of N Interval Length

SPEAKER (f_S)	VALUE OF PARAMETERS					
	A	α	n	H_{max}	μ	$P(\mu)$
M1(8 kHz)	1	0.51	13	0.021	0.075	≈1
M2(8 kHz)	1	0.48	13	0.038	0.140	≈1
M3(8 kHz)	1	0.41	12	0.024	0.083	≈1
W1(8 kHz)	1	0.74	12	0.045	0.160	≈1
W2(8 kHz)	1	0.66	15	0.060	0.230	≈1
M4(32 kHz)	1	0.56	13	0.040	0.140	≈1
W3(32 kHz)	1	0.52	14	0.055	0.200	≈1
AVERAGE VALUES	1	0.554	13.1	0.0404	0.1469	≈1

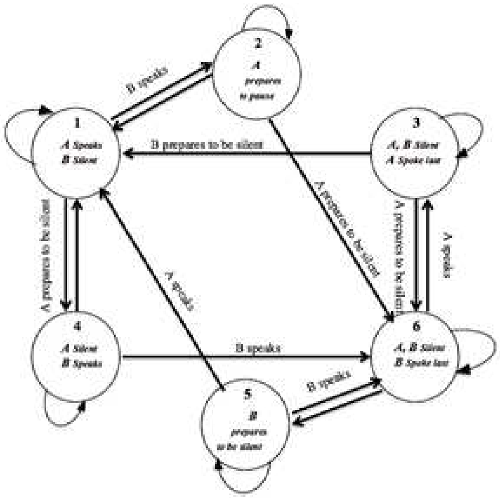

Figure 1. Graph of the Six-state Brody model

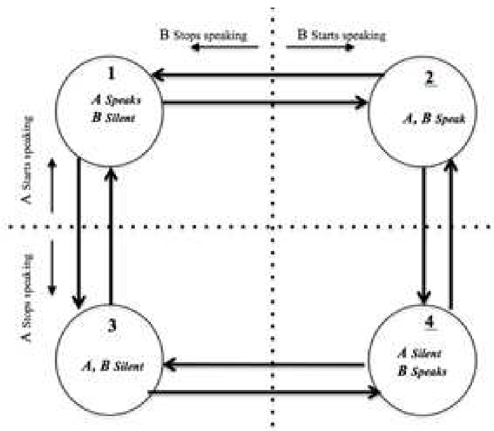

Figure. 2. Graph of four-state dialogue model

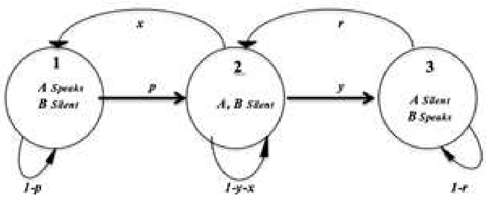

Figure. 3. Graph of three-state dialogue model

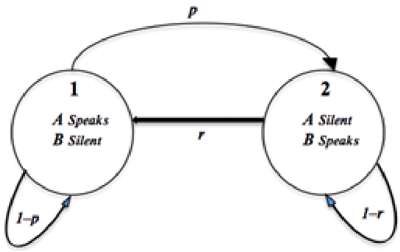

Figure. 4. Graph of Two-state dialogue model

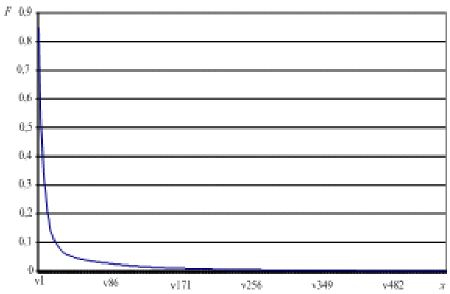

Figure. 5. *PDF* of the duration of active packets for all conversations

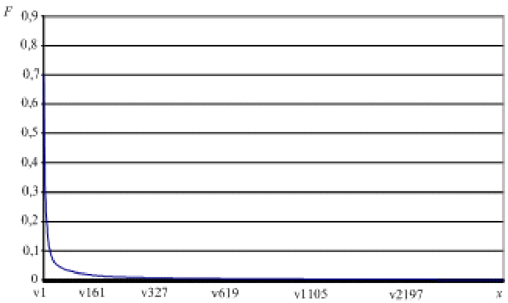

Figure. 6. PDF of the duration of *passive packets* for all conversations

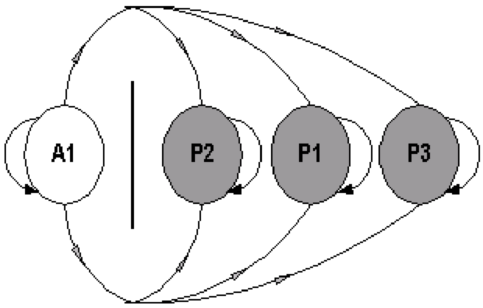

Figure. 8. Graph of the *four-state* model

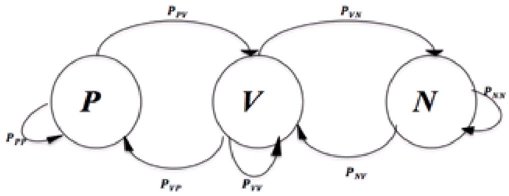

Figure. 9. Graph of a 3–state transition Markov model of monologue speech

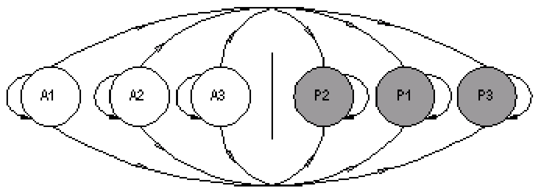

Figure. 7. Simplified graph of the *six-state* model